**Using machine learning techniques to analyse economic statistics: a case study with HMRC Trade in Goods statistics**

Andy Banks, Office for National Statistics

Recent developments in machine learning and data visualisation techniques have led to growing interest in their application to statistical and econometric analysis.

This paper examines two potential applications of machine learning in analysing economic statistics: its use in regression analyses and outlier detection. The paper applies both to a HMRC administrative dataset that measures UK exports of products to the rest of the world.

First, we consider the use of wide and deep neural networks to produce linear regression models. These techniques are tested on the administrative data to produce predictive models of UK exports to other countries, and are tested against more ‘general’ regression techniques.

Secondly, we test unsupervised machine learning algorithms to show their effectiveness in measuring outliers in trade administrative data. This can be used to communicate the impact that a small number of outliers may have on headline aggregate statistics.

Finally, the advantages and limitations of a range data mining techniques are discussed.

Key Words: trade statistics, linear regression, outlier detection, unsupervised machine learning, neural networks

Andy Banks

2.201, Office for National Statistics, Newport

Telephone: 01633456486

Email: Andrew.Banks@ons.gsi.gov.uk